# Higher-Level, Downward and Specific Causation

Max Kistler

**Higher-Level, Downward and Specific Causation**

Max Kistler, Université Paris 1 Panthéon Sorbonne
IHPST (Institut d'histoire et de philosophie des sciences et des techniques), Paris
mkistler@univ-paris1.fr

Forthcoming in M. Paolini Paoletti, F. Orilia (eds.), *Philosophical and Scientific Perspectives on Downward Causation*, Routledge, 2017. Please quote only from published version.

**Abstract**

The Interventionist account of causation (Woodward 2003) seems to provide a rigorous framework for evaluating the possibility of downward causation. However, it has turned out 1) that only a modified version (Woodward 2014) of interventionism can be applied to situations of apparent downward causation and that 2) this model, though compatible with downward causation, makes it in principle impossible to find empirical support for downward causation (Baumgartner 2013). In this paper I show in which sense downward causation can be justified by using more fine-grained notions of causation, such as stable, proportional and specific causation (Woodward 2010). In particular, the intervention on a higher-level variable H($t$) with respect to a lower-level variable P($t$*) (where $t$* is later than $t$) may be more proportional compared to the parallel intervention on lower-level variable R($t$) w.r.t. to P($t$*), if R($t$) is too determinate with respect to P($t$*), i.e. if an intervention on R($t$) is not necessary for manipulating P($t$*).

Physics is not the only science. If, as physicalism assumes, everything is exclusively composed of parts that are objects of fundamental physical theories, why isn't there only fundamental physics? More specifically, with respect to the search for causes, is it conceivable that special sciences identify *causal relations* holding among events that involve objects and properties that are not in the scope of fundamental physics? Can there be causal explanations that make reference to causal relations invisible for physics? In other words, can special sciences such as solid state physics, chemistry, biology or psychology identify their own causes, or do all these sciences have a merely heuristic status, in the sense that they never identify the real causes but only some form of useful simplification?

The world seems to be full of cases of causation where both the cause and the effect are *higher-level* in the sense that they are properties that are not directly the object of physics. Perception of a predator by a prey $M_1(t)$ causes flight $M_2(t')$. Both perception and flight are "higher-level" concepts, in the sense that they characterize living organisms, which are complex systems. I will use the notion of a level in the sense in which a property that characterizes a complex system is at a higher level than properties characterizing components of these complex systems (Craver 2007, p. 188).

*Downward* causation seems just as ubiquitous. The fact that the prey's perception $M_1(t)$ of a predator at time t causes the contraction of a muscle $P_2(t')$ of the prey at time t', where $P_2(t')$ is a component of flight $M_2(t')$, seems to be straightforward case of true downward causal statement. In general, I will speak of "higher-level causation" when a higher-level variable $M_1$ influences a higher-level variable $M_2$, and of "downward causation" when a higher level variable $M_1$ influences a lower-level variable $P_2$.

Here is a case of downward causation in physics. Let us consider the phase transition of a particular piece of Nickel from the ferromagnetic to the paramagnetic

state. The phase transition results from heating, which raises its temperature above the critical temperature $T_c$. In the ferromagnetic state, the piece is a macroscopic magnet with an overall magnetic moment because the spins of the atoms composing it are all aligned. When its temperature rises above $T_c$, this alignment disappears. In the resulting paramagnetic state the overall magnetic moment is zero. Moreover the mean value of the spin of any small set of neighboring atoms within the piece of nickel also becomes zero, whereas it had some positive value s in the ferromagnetic state. It would seem that the macroscopic property of being at a temperature T above $T_c$ at t is the cause of the microscopic effect that the mean value of the spin of atom i and its nearest neighbors is 0 at t'. Downward causation seems to be common also in other sciences. In biology, the stimulation of a neuron influences the state of ionic channels in its membrane. The former characterizes a change in a property lying at a higher level (that of the neuron) than the latter (the channels are constitutive parts of the neuron). In psychotherapy, changing a patient's beliefs can alleviate depression (Cuijpers et al. 2008), and modify its neural basis: It can modify abnormal regional cerebral blood flow and glucose metabolism (Kalia 2005). A change in belief is a change in a property characterizing the whole person whereas the effect is a change at a lower level concerning a property of a part of the person's brain.

Partisans of physicalism are attracted to the view that all causes are physical. According to such a view, causal statements of special sciences that mention non-physical properties are mere simplified ways of speaking, which are ultimately made true by physics. It seems even more obvious to most physicalists that properties that are not themselves physical can never causally influence physical properties: there cannot be such a thing as downward causation.

However, the reasoning behind these physicalist theses relies on metaphysical principles that are not themselves directly justified by science. For one, Kim's (1998) argument for the impossibility of both higher-level causation and downward causation relies, e.g., on the "principle of the causal closure of the physical domain". One may question the monopoly of physics in the identification of causes and thereby justify the competence of special sciences in identifying causes by challenging such metaphysical principles. However, a more straightforward way consists in justifying the existence of higher-level causes directly by the existence of successful higher-level sciences.

In this paper I will examine whether and to what extent the interventionist account of causation (Woodward 2003) can contribute to justify higher-level and downward causation. This question has given rise to a controversy in the literature, some authors arguing that it can (Shapiro and Sober 2007, Raatikainen 2010) and others that it cannot (Baumgartner 2009, 2010, 2013; Marcellesi 2010). I do not ask whether the interventionist framework all by itself provides the conceptual tools for justifying higher-level and downward causation, but whether it can be used as a complement to an analysis of causation in terms of transference (Kistler 2006; 2013). In such a framework, causation is taken to be a relation between localized events, i.e. contents of local regions of space-time, which holds because some amount of some conserved quantity such as energy is transmitted between them. However, to provide a satisfactory causal explanation, it is often not enough to identify and make reference to the events that are related and the simple fact that they are causally related. A causal explanation is supposed to provide information about which property of the cause event was responsible of a given property of the effect event (Kistler 2014). The issues of higher-level and downward causation bear on the question of whether higher-level properties can play the role of such causally responsible properties, with respect to a

given property of the effect event. In this paper, I will address the question as to whether the interventionist framework of analysing causation as a relation between variables can be used to supplement the transference account, in the sense of determining which of the properties of a cause event is responsible of a given property of the effect event.

Let me illustrate with the example of the phase transition in Nickel. Let the cause be the event at which a given macroscopic piece of Nickel that is in a ferromagnetic state and at a temperature below $T_c$ absorbs at t a certain amount of heat, and let the effect event be a microscopic portion of the piece of Nickel containing atom i and its nearest neighbors, a little later at t' when the metal's temperature has risen above $T_c$. The question is whether the interventionist account can be used to justify the statement that the piece of Nickel's temperature raising above $T_c$ at time t is causally responsible for the fact that the mean value of the spin of the atoms neighbouring atom i becomes zero at t' a little later.

The interventionist account of causation seems to provide the means for justifying this judgment. The intervention that modifies the macroscopic cause variable temperature T by warming the metal up, thereby switching the value of T from $T_1<T_c$ to $T_2>T_c$, is followed by a change in the microscopic effect variable $S_i$ that represents the mean value of the spin of atom i and its nearest neighbors: The value of $S_i$ switches from s (corresponding to the mean value of atomic spin in the ferromagnetic state) to 0.

## Interventionist analysis of causation

The interventionist analysis of causation makes explicit the experimental strategy used in science for discovering causal relations among variables. It is not intended to provide an analysis of causation as a relation between individual spatio-temporally localized events, but an analysis of causation as a relation among properties of events[1], which can be represented by variables.

The fundamental idea of this approach is this. One variable X causally influences a second variable Y if and only if there are interventions (satisfying certain conditions) such that modifying the value of X by such an intervention also modifies the value of Y. In Woodward's terms, "X causes Y if and only if there are background circumstances B such that, if some (single) intervention that changes the value of X (and no other variable) were to occur in B, then Y or the probability distribution of Y would change" (Woodward 2010, p. 290).

The interventionist conditions for the existence of a causal relation between variables X and Y correspond to experimental and observational criteria on which scientific method grounds the judgment that X causally influences Y. The general idea of the recipe is this. Find a variable I, corresponding to a possible way of modifying the value of the cause variable X, which satisfies the following conditions for being an intervention variable on X with respect to Y (Woodward 2003, p. 98).

(IV)

1. I directly influences X but does neither directly influence Y nor any other variables influencing Y that do not lie on the causal path from I to X to Y.

2. I completely "controls" X, in the sense that the intervention I cuts off all other influences on X.

3. The intervention I has an origin independent of the variables that are being investigated. In particular, I is not statistically correlated with any causes of Y that do not lie on the causal path from I to X to Y.

[1] Events are here conceived as particulars, i.e. as what fills a given space-time zone.

Then manipulate X by way of I and observe whether changes in X are accompanied by changes in Y. If and only if they are, X causally influences Y.

In the original framework of interventionism (Woodward 2003) it is impossible to justify causal judgments in which a higher-level variable X acts as a cause of a lower-level variable Y, as soon as lower-level variables SB(X) in the supervenience base of X are also taken into account (Baumgartner 2009, 2010, 2013; Marcellesi 2010). This leaves open the possibility to justify that X causes Y by simply not taking into account any lower-level variables SB(X) on which X supervenes. However, such a justification would be ad hoc, given that the main challenge consists in justifying the causal role of X, against the claim that all causes of Y lie at the same level as Y, i.e. at the level of the variables SB(X) in the supervenience base of X. Moreover, even if the omission of variables SB(X) might make it possible to provide a formal justification of a downward causal claim X→Y, higher-level variables could never be causes in situations where variables in their supervenience base are also causes. Thus, such a justification would exclude by stipulation the possibility that *both* SB(X) and X causally influence Y.

Shapiro and Sober (2007) and Woodward (2014) have suggested to modify the interventionist framework so as make it possible to justify causal statements according to which supervenient variables are causes without excluding variables in the supervenience base from consideration. Such a modification opens up the possibility to use the interventionist framework to argue against eliminativism and epiphenomenalism with respect to higher level variables.

Both the conditions (IV) on intervention variables and the definition of direct causation must be modified with respect to Woodward's (2003) original analysis. The leading idea for the modification of (IV) is that the variables SB(X) in the supervenience base of the cause variable X should be excluded from the set of variables that must be held fixed during an intervention in X. "To assess whether X causes Y, the common causes of X and Y must be held fixed, but not the microsupervenience base of X" (Shapiro & Sober 2007: 8). For it is not only impossible by definition of supervenience to hold variables SB(X) fixed during an intervention on X, but such a requirement does not correspond to scientific standards of experimental control of causal hypotheses. "It is inappropriate to control for supervenience bases in assessing the causal efficacy of supervening properties" (Woodward 2014, p. 21).

In the framework that results from the modification of (IV) along these lines – let us call it (IV*) - a variable I may count as an intervention on X with respect to Y even though every change in the value of I that changes the value of X also necessarily changes the value of SB(X), as sketched in fig. 1.

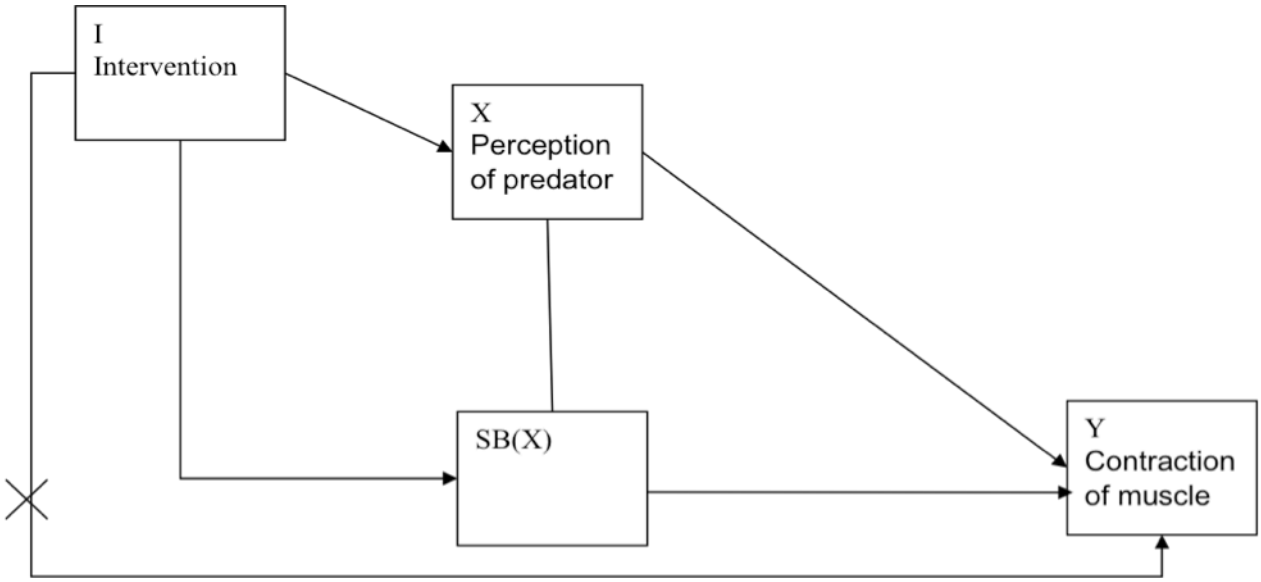


Fig. 1: Sketch of an intervention by I on X, which is also an intervention on SB(X). The cross represents the rule that for the variable I to be an acceptable intervention variable, it must not directly influence Y. There is no cross on the arrows I→X and I→SB(X), which represents the fact that I may influence both X and SB(X).

In the same spirit, the conditions for a variable X to be a direct cause of variable Y can be weakened in the following way, so that it becomes conceivable that a higher-level cause X is a direct cause of Y (which may be at the same level as X or at a lower level)

(M*) A necessary and sufficient condition for X to be a (type-level) direct cause of Y with respect to a variable set V is that there be a possible intervention on X that will change Y or the probability distribution of Y when one holds fixed at some value all other variables $Z_i$ in V, with the exception of the variables in the supervenience base of X and of Y (if V contains such variables).

There has been a controversy over whether these new definitions determine the conditions for X to be a direct cause of Y in such a way as to distinguish them from the conditions under which it is rather SB(X) that causes Y. In case Y is a variable at the level of SB(X) the question is whether these conditions make downward causation (X→Y) empirically distinguishable from lower-level causation (SB(X)→Y).

Before I answer this question on downward causation, let me consider the question whether (IV*) and (M*) make the higher-level causal claim that X causes Y empirically distinguishable from the corresponding lower-level claim that SB(X) causes SB(Y).

It seems to be conceivable that there are situations of both following types:

1) Situations (sketched in fig. 2, following Woodward, forthcoming, p. 10) containing two higher-level variables $M_1$ and $M_2$, supervening respectively on variables $N_1$ and $N_2$, where there is causal influence at *both levels*, i.e. $N_1$ influences $N_2$ and $M_1$ influences $M_2$.

2) Situations (sketched in fig. 3) containing two higher-level variables $M_1$ and $M_2$ that are *not* causally related but which supervene on variables $N_1$ and $N_2$ which *are* so related.

If both situations are conceivable and empirically different, the statement that $M_1$ causes $M_2$ has an empirical content that is independent from the statement that $N_1$ causes $N_2$. The fact that $N_1$ causes $N_2$ leaves it open whether or not $M_1$ also causes $M_2$.

However, it has been questioned whether the objective difference between these two kinds of situation is sufficient to justify the claim that the modified interventionist framework provides verification conditions, and thus gives empirical content, to higher-level causal claims (Baumgartner and Gebharter 2015). The problem is that there seems to be no *sufficient* empirical condition that would establish that a given situation is one where there is causation at both lower and higher levels (as in fig. 2).

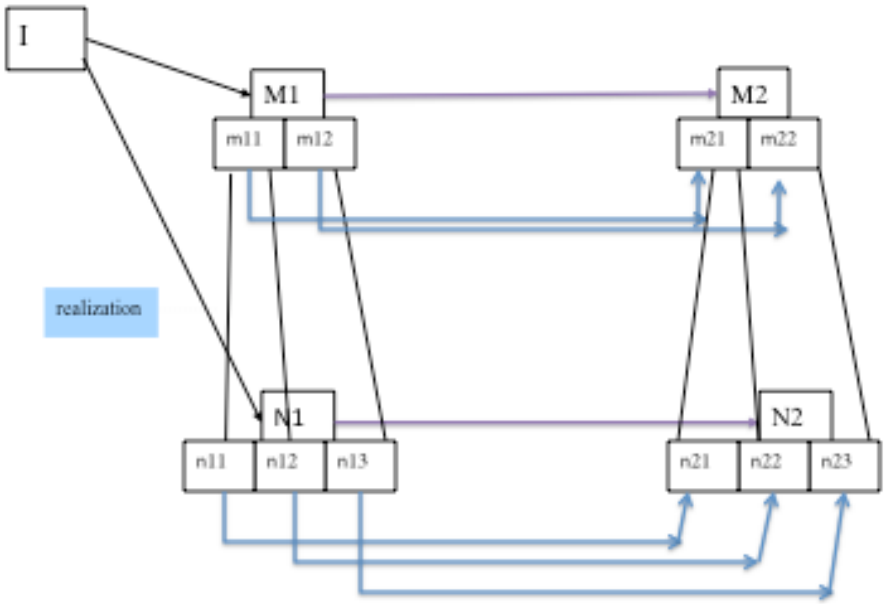


fig. 2. Model of a situation in which there is *both* lower-level causal influence $N_1 \rightarrow N_2$ and higher-level influence

$M_1 \rightarrow M_2$.

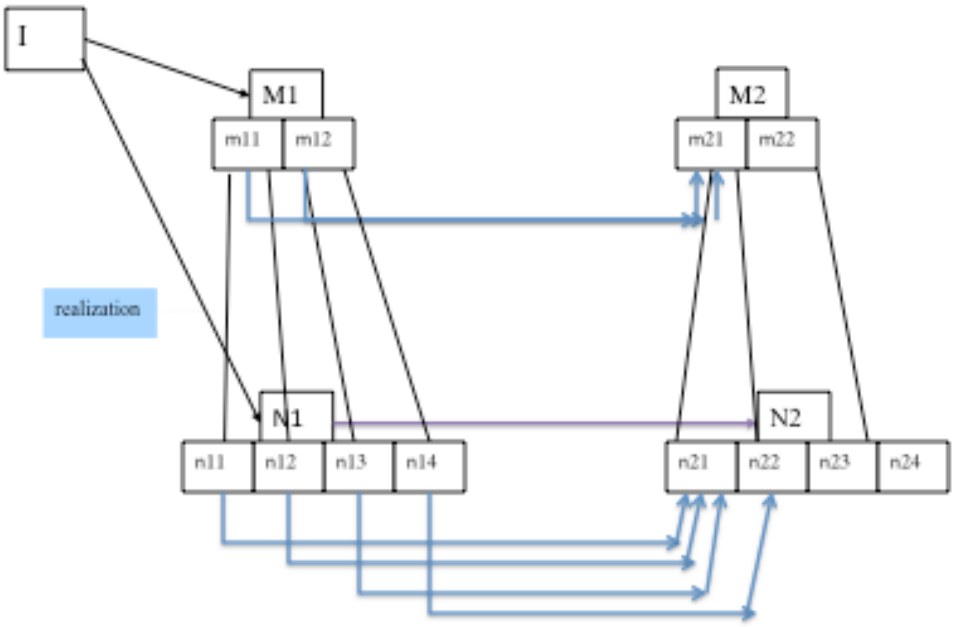


fig. 3. Model of a situation in which there is lower-level causal influence $N_1 \rightarrow N_2$ *but no* parallel higher-level influence $M_1 \rightarrow M_2$.

Let me explain. $M_1$ and $M_2$ are causally related in the framework of (M*) iff there is at least one possible change in the values of $M_1$ (brought about by an intervention) that would change the value of $M_2$. And $M_1$ and $M_2$ are *not* causally related in the framework of (M*) iff there is *no possible change* in the values of $M_1$ (brought about by an intervention) that would change the value of $M_2$. Fig. 2 illustrates the former, fig. 3 the latter.

Both figures 2 and 3 represent possible situations containing higher–level variables $M_1$ and $M_2$, and variables $N_1$ and $N_2$ in their respective supervenience bases, where $N_1$, the supervenience basis of $M_1$, exercises a causal influence on $N_2$, the supervenience basis of $M_2$. The comparison of the two situations sketched in fig. 2 and 3 shows that the higher-level influence $M_1 \rightarrow M_2$ can be experimentally distinguished from the lower-level influence $N_1 \rightarrow N_2$. One causal relation can exist without the other. This shouldn't be so surprising, given that the concept of supervenience is mostly used in situations in which it is asymmetric, i.e. in which changes in the supervenient variables are always accompanied by changes in the supervenience base, but in which the reverse does not hold, i.e. where some changes at the level of variables in the supervenience base are not mirrored by any changes and causal influences at the level of the supervenient variables. This is the case when supervenience is used to characterize the relation between psychological properties and neurophysiological properties: the former are supposed to supervene on the latter but not the reverse. One psychological property can correspond to many underlying neurophysiological properties, whereas

only one psychological property is compatible with any given neurophysiological property.

What is special in the case sketched in fig. 3 with respect to usual situations of supervenience, is that not only are some particular interventions at the level of $N_1$ that cause changes in $N_2$ not mirrored by parallel changes at the level of the supervenient variables (and thus, some causal influences at the level at the level of the supervenience basis are not mirrored by causal influences at the level of the supervenient variables), but that there is *no causal relation* at the higher-level *between the variables themselves*. This means that it is objectively *impossible* to influence $M_2$ by intervening on $M_1$, i.e. by modifying the value of $M_1$.

The problem is that there is no empirical criterion that could justify the judgment that a given situation is of the type represented in fig. 3, i.e. of a sort in which it is *impossible* to modify $M_2$ by intervening on $M_1$. One can justify that it is possible to modify $M_2$ by intervening on $M_1$, simply by doing it. But no finite set of observations can guarantee that it is impossible to modify $M_2$ by intervening on $M_1$, and in particular, it is not sufficient to show that so far, no intervention on $M_1$ has modified $M_2$.

So can the causal influence of supervenient variables be assessed (by interventionist means) independently from the assessment of the causal influence of variables in their respective supervenience bases, as Woodward (2008, 2008b, 2014, forthcoming) and Menzies and List (2010) claim? In other words, can it be justified on empirical grounds that a situation is of the type sketched in fig. 2 rather than of the type sketched in fig. 3? The answer is that it can, but that the fact that the situation corresponds to fig. 3 may in some cases be established only on inductive grounds (Baumgartner and Gebharter 2015). This is the case if not all possible values of $M_1$ and $M_2$ are known and also if the dependence of $M_2$ on $M_1$ is probabilistic rather than deterministic.

In such situations, single experimental manipulations can only establish that $M_1$ causally influences $M_2$ (because they can establish that some changes in the value of $M_1$ are followed by a change in the value of $M_2$, by provoking such changes in the value of $M_1$). However, if one does not know all the possible values of $M_1$ or if the influence of $M_1$ on $M_2$ is probabilistic, neither single manipulations nor finite series of such manifestations can establish that $M_1$ *does not* influence $M_2$, i.e. that there can be no change in the values of $M_1$ that would be followed by a change in $M_2$.

With respect to downward causation, Baumgartner (2010; 2013) has argued that an interventionist account based on conditions (IV*) and (M*) does not provide a framework that would allow empirical justification of downward causation. In that account, relations of causal influence remain "underdetermined" (between downward and same-level causal influence) because it yields the result that two causal statements – that X directly causes Y and that SB(X) directly causes Y – are true under the same conditions, so that the analysis violates the interventionist maxim according to which different causal claims must be justified by different relations of manipulation.
Here is Baumgartner's argument: If $M_1$ is a higher-level variable, $P_1$ a variable characterizing its supervenience base, then the statement according to which $M_1$ causes $P_2$ (which may be at the level of the supervenience base $P_1$), as sketched in fig. 4, and the statement according to which it is rather $P_1$ that causes $P_2$ (as sketched in fig. 5) are "empirically indistinguishable" (Baumgartner 2010, p. 19; 2013, p. 22).

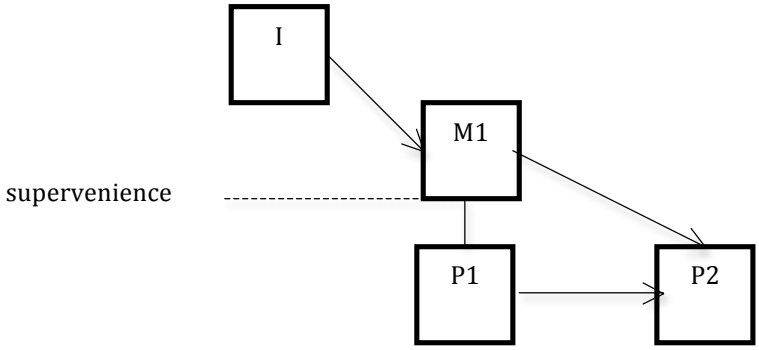


Fig. 4. Intervention on higher-level variable with downward causation

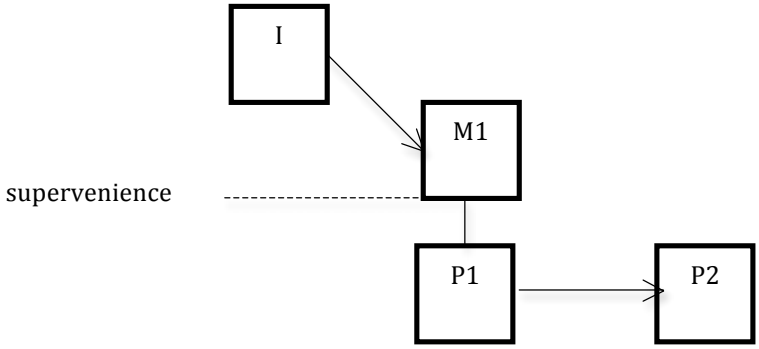


Fig. 5. Intervention on higher-level variable *without* downward causation

"The epiphenomenalist structure" sketched in figure 5 "generates the exact same difference-making relations or correlations under possible interventions as" (Baumgartner 2013, p. 21/2) the structure sketched in figure 4, in which variable $M_1$ exercises downward causal influence on $P_2$. However, it is not true that both statements have the same empirical truth-conditions. Just as for higher-level causal judgments, the empirical content of a downward causal statement differs from the content of the corresponding lower-level causal statement. Here is a sketch of the formal structure of two situations in which there is causal influence between two lower-level variables $N_1 \rightarrow N_2$. In the first (sketched in fig. 6), there is also downward causation $M_1 \rightarrow N_2$, whereas there is no such downward causal influence in the second (sketched in fig. 7). The very conceivability of the second situation shows that a downward judgment such as $M_1 \rightarrow N_2$ has empirical content.

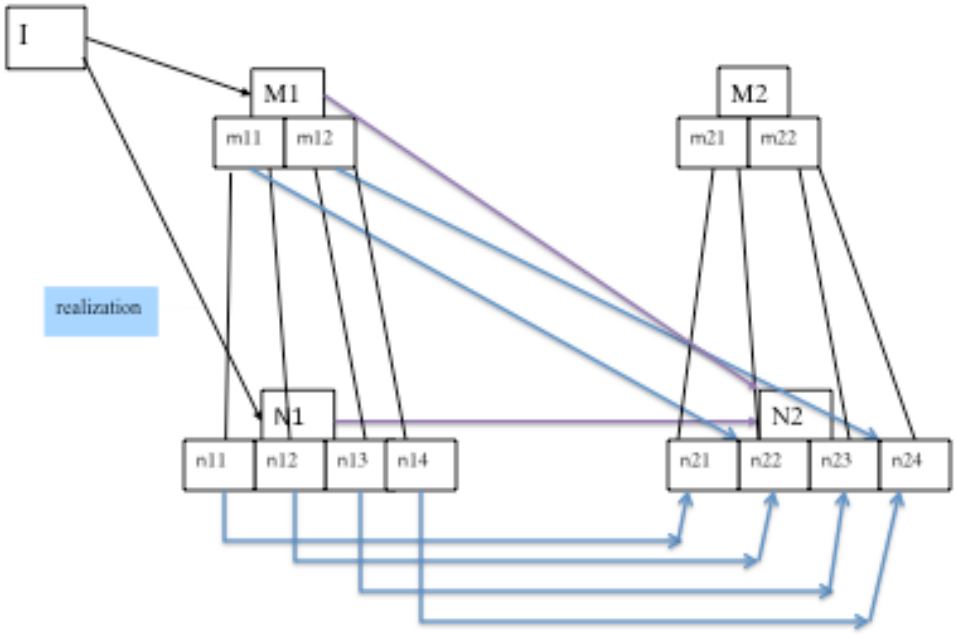


Fig. 6. Model of downward causation with parallel lower level causation.

In situations that have the structure of fig. 6, there is lower-level causation because interventions on $N_1$ can make a difference to the value of $N_2$, but there is also downward causation because interventions on $M_1$ can change the value of $N_2$: a switch shifting the value of $M_1$ from $m_1$ to $m_2$ brings about a switch of the value of $N_2$, from (either $n_{21}$ or $n_{22}$) to (either $n_{23}$ or $n_{24}$).

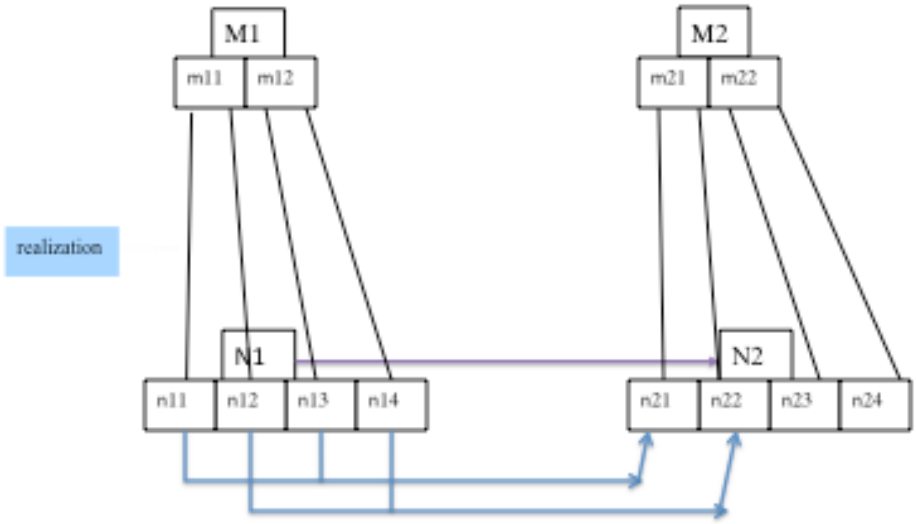


Fig. 7. Model of lower-level causation, without downward causation.

However, the fact that $N_1$ influences $N_2$ at the lower level does not by itself entail that there is also downward causal influence from $M_1$ on $N_2$. This is shown by the existence of situations that have the structure of fig. 7. In such situations, there is lower-level causation $N_1 \rightarrow N_2$ because some interventions (such as a switch from $n_{11}$ to $n_{12}$) change the value of $N_2$ (from $n_{21}$ to $n_{22}$). But there is *no downward* causal influence $M_1 \rightarrow N_2$ because no switch in the value of $M_1$ induces any reliable switch in the value of $N_2$. Each of the values of $M_1$ ($m_1$ and $m_2$) can yield $n_{21}$ and each can yield $n_{22}$, so that the difference between $n_{21}$ and $n_{22}$ does not correspond to any difference between different values of $M_1$.

Here are two situations that have the structure of fig. 6 and 7. Let $M_1$ represent the color of a traffic light, with $m_{11}$ being the value for green, and $m_{12}$ for red. Let $M_2$ represent the state of a car passing the traffic light, with $m_{21}$ being the value for the car moving and $m_{22}$ for the car stopping.

Let $N_1$ represent the state of the electric circuit in the traffic light, where $n_{11}$ and $n_{12}$ are two states where current flows through the green lamp, and $n_{13}$ and $n_{14}$ states where current flows through the red lamp. Moreover $n_{11}$ and $n_{13}$ also activate a sound for blind people, something neither $n_{12}$ nor $n_{14}$ do. $N_2$ represents the state of the engine of the car: values $n_{21}$ and $n_{22}$ represent states where it makes the car move, where $n_{21}$ makes the car move in automatic mode.

If the driver respects the rules, the situation that has the structure of fig. 6: There is downward causal influence from the color of the traffic light to the motion of the car: green light makes the car move ($n_{21}$ or $n_{22}$), whereas red light makes it stop ($n_{23}$ or $n_{24}$).

If the driver is colorblind or inattentive, the situation may have the structure of fig. 7: Both states of the traffic light make the car move. But let us furthermore suppose that, to compensate for the driver's distraction or poor discrimination of colors, the car has a mechanism that puts the engine in automatic mode if and only if it receives the sound emitted by a traffic light. Then there is no downward causation: the color of the traffic light makes no difference to the state of motion of the car. However, there is lower-level causation (just as in the situation corresponding to fig. 6): With the colorblind driver, the difference between states of the traffic light that produce a sound ($n_{11}$ and $n_{13}$) and those that do not ($n_{12}$ and $n_{14}$) makes a difference to the state of the engine of the car, between the automatic and the non automatic mode.

The existence of these two types of situation, sketched in fig. 6 and 7, shows that the statement of downward causal influence $M_1 \rightarrow N_2$ has its own specific empirical content, distinct from the statement of lower-level causal influence $N_1 \rightarrow N_2$. For the same reason as in the case of higher-level causal statements, it can be difficult to find out whether there is *no downward* causal influence. In certain situations, the absence of downward causation can be justified only inductively (lower-level causation being presupposed). This is the case if either not all values of $M_1$ are known or if the causal influence $M_1 \rightarrow N_2$ is probabilistic. In such circumstances, it can be the case that no downward influence has been observed although it objectively exists.

To sum up, supervenience guarantees that there can be neither higher-level causation nor downward causation without lower-level causation. However, there is no "upward exclusion": The presence of causal influence at some level (e.g. physical) $N_1 \rightarrow N_2$ leaves the question open whether there is also higher-level causal influence between variables that supervene on $N_1$ and $N_2$, and whether there is downward causation $M_1 \rightarrow N_2$ or not. Given $N_1 \rightarrow N_2$, there may be and there may not be higher-level influence $M_1 \rightarrow M_2$, and there may be, or there may not be, downward influence $M_1 \rightarrow N_2$. The difference between situations where there is and where there is not higher-level or downward influence has empirical content because it corresponds to different patters of difference-making.

**Specific causation**

Supervenience guarantees that there cannot be "downward exclusion": Higher-level and downward causation is always accompanied by physical level causation. This seems to be in straightforward contradiction with List and Menzies' (2009) thesis that there can be "downward exclusion" in the sense that there are situations in which there is higher-level causation or downward causation without any underlying physical level causation. According to List and Menzies, this is possible if the higher-level cause is "realization-insensitive".

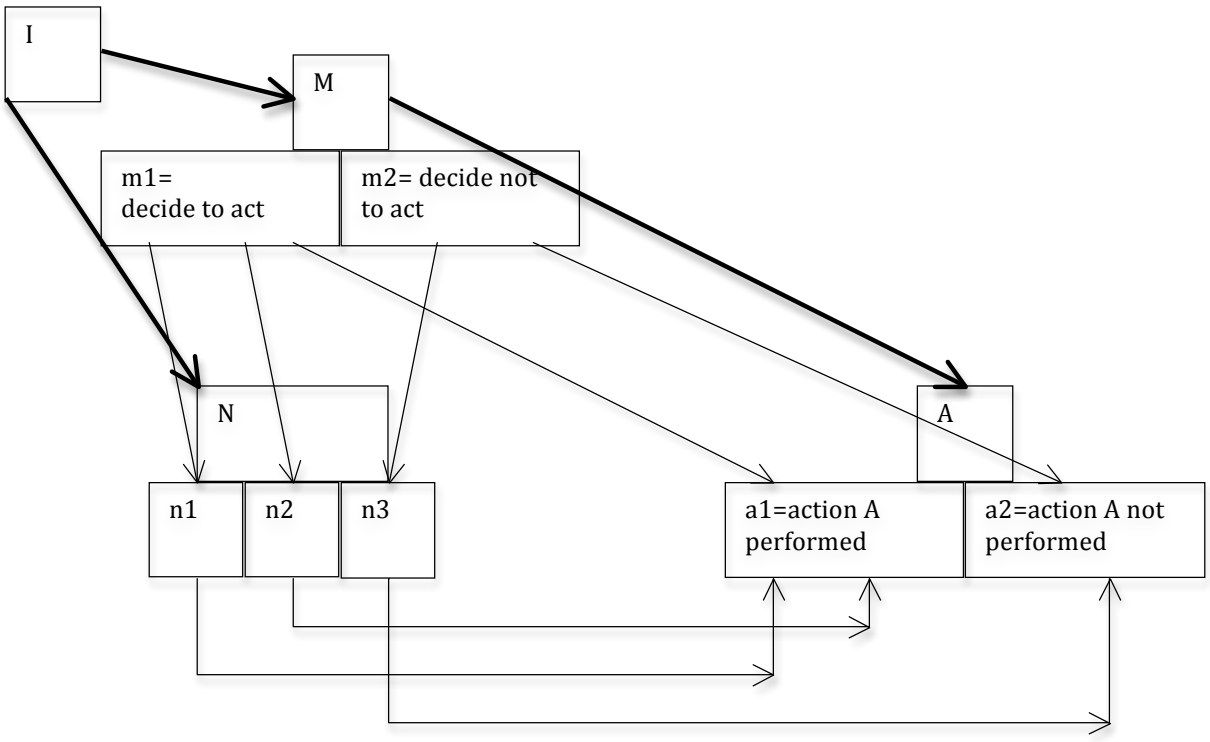


Fig. 8. Structure of a situation in which downward causal influence M→A "excludes" lower-level influence $N_1$→A. Thick arrows represent causal influence between variables. Thin arrows indicate which value of the cause value gives rise to which value of the effect variable.

Fig. 8 represents the structure of a situation of "downward exclusion" sketched by List and Menzies (2009). They argue that there can be downward causal influence M→A of a higher-level variable M on a lower-level variable A without any influence on A of the supervenience basis $N_1$ of M, because M but not $N_1$ is a "difference-maker" with respect to A.

In a situation of the structure sketched in fig. 8, M is a difference-maker for A because switching from one value of M to another makes for a switch in the value of A. However, $N_1$ is *not* a difference-maker for A in List and Menzies' sense because *not every* switch in the value of $N_1$ makes for a switch in the value of A. More precisely, *some* switches in the value of $N_1$ cause a switch in the value of A, *but not all*. Some switches, such as the switch from $n_{11}$ to $n_{12}$ make no difference to the value of A, which is $a_1$ in both cases. In other words, $N_1$ is no difference-maker for A in the sense in which List and Menzies (2009) use this concept, because some values of $N_1$ (such as $n_{11}$ and $n_{12}$) are not necessary for the value of A they are associated with. Given that M is and $N_1$ is not a difference-maker for A, there is downward exclusion with respect to difference-making: the higher-level M "excludes" the lower-level variable $N_1$ from being a cause of A, in the sense of being a difference-maker.

I have argued above that higher-level and downward causation are always accompanied by parallel lower-level causation. This thesis appears to be incompatible with Menzies and List's thesis that there can be downward exclusion. But the appearance of incompatibility dissolves once the difference between causation and difference-making in List and Menzies' sense is taken into account.

List and Menzies' concept of difference-making is similar to what others (Yablo 1992, 1997; Woodward 2010) have called "proportional causation". Let me introduce this concept with Yablo's (1992, p. 257) example of a pigeon that has been conditioned to peck at all and only red targets. It reliably pecks at all targets of all shades of red but not at any targets of any other colour. Now consider a type of situation in which a scarlet target is shown and in which the perception of that scarlet target causes the pigeon to peck. In the interventionist framework introduced above, both the variable R corresponding to the perception of a red target and the variable S corresponding to the perception of a scarlet target are causes of the variable P corresponding to the pigeon's pecking. Both of the following causal statements are true.

(1) *perception of scarlet* target causes the pigeon to peck.

(2) *perception of red* target causes the pigeon to peck.

However, (2) seems intuitively "more correct" than (1). According to Woodward (2010, p. 298), the second judgment "furnishes a better explanation" than (1). The difference between (1) and (2) can be explained in terms of the proportionality of the cause variable R ("perception of red") with respect to the effect variable P ("pigeon pecking"). A causal statement with R as the cause variable does not mention any irrelevant detail, as the choice of the variable S ("perception of scarlet") does, but R is sufficiently determinate to be related to the effect variable deterministically, or at least with a stable high probability, >> 0.5. This is not the case if the cause is represented by the variable W corresponding to the perception of a warm colour, where "warm colour" means "a colour in the part of the spectrum lying between red and yellow".

I propose to call the relevant concept "specificity", and leave proportionality for the mathematical relation between two variables X and Y, where one is the product of the other with a scalar factor $c$, $Y=cX$. Here is a proposal to define the notion of C being a specific cause of E for *many-valued* variables.

Let C and E be variables with many values, either on a continuous or on a discrete spectrum.

(S-P) C is a *specific* cause of E iff

1) C causes E (in the sense that *some* interventions on C, changing the value of C, change the value of E) and

2) C represents a natural property F of the cause event and E represents a natural property G of the effect event such that both C and E can vary within a continuous or many-valued discrete spectrum, values ci of C correspond to the values of the property F, values $e_i$ of E correspond to the values of G, and

3) The values of E are an *injective function* of the values of C.

A function is said to be injective iff for all pairs of values $x_i \neq x_j$, $f(x_i) \neq f(x_j)$, or in other words, there is no pair of values $x_i \neq x_j$ with $f(x_i)=f(x_j)$. As an illustration, take a domestic dimmer, which allows modifying the intensity of light among values on a continuous spectrum by rotating a switch. One simple mechanism implementing such a controlling device is with a coil-rotation transformer. Such a mechanism implements a causal chain, from the position of the switch (P) to the voltage (V) to the intensity of light (L). Let us suppose, for simplicity's sake, that the values of all three variables are proportional to each other in the mathematical sense. Let us consider only the relation between V and L.

The concept of a specific cause can be used to choose an interventionist variable that represents the property that is causally responsible for the intensity of light L. If the values of V represent the values of the voltage, say within the interval (0, 220V) and the values of L the values of the intensity of light, say between the interval (0, 1000 lumen), then V is a cause of L in the interventionist sense and a specific cause in the sense of (S-P): V is a *cause* of L because *some* interventions (turning the switch) that bring about changes in the value of V are followed by a change in the value if L. V is *specific* for L because the values of L are an injective function of the value of V. Every modification of the value of V corresponds to a change in the value of L[2].

**Differences in specificity relative to the level of properties within a mechanistic system**

To bring the concept of specificity to bear on the issue of higher-level and downward causation, we need to show how it can be used to compare variables characterizing a system by representing properties at different levels. In this sense, variables characterizing a whole system are at a higher level than variables characterizing functional parts of the system. I will now show how the concept of specificity can be applied to variables describing systems at different levels in this sense.

Let me introduce this use of the notion of specificity with an example from neurophysiology (Kandel and Siegelbaum 2000). Let us look at the cause of the post-synaptic potential (PSP) in a muscle cell, which itself typically triggers an action potential in the muscle cell. The PSP is the result of a chain of events and processes. Let us look at the step in the chain where many vesicles containing acetylcholine (Ach, the messenger substance) fuse with the plasma membrane of the motor neuron. This fusion results in release of the Ach contained in the vesicles into the synaptic cleft.

Let us compare the following two variables, one at the level of the whole neuron and one at the lower level of components of the neuron: M, the higher-level variable, represents the quantity of Ach released into the synaptic cleft. Let M be a 2-valued variable: value $m+$ corresponding to a quantity larger than the threshold for triggering an action potential in the muscle cell, $m-$ corresponding to values between 0 and the threshold. The effect variable A, representing the triggering of the PSP in the muscle cell, is also a 2-valued variable: value $a+$ represents triggering of a PSP, value $a-$ represents absence of triggering.

Let us consider the following lower-level variable P. P characterizes the fusion of the individual vesicles with the plasma membrane. It is convenient to construe P as a vector, with one component $p_i$ for each vesicle. Each component has a value $p_i+$ representing the fact that vesicle $i$ merges and $p_i-$ the fact that it doesn't.

Both P and M are causes of A. P is a cause of A because there are interventions on P that switch its value, e.g. interventions that make all vesicles merge with the membrane, switching the value of P from ($p_1-$, $p_2-$, …., $p_n-$) to ($p_1+$, $p_2+$, …. $p_n+$), which make the value of A switch from $a-$ (no PSP) to $a+$ (PSP). M is also a cause of E because there are interventions on M that switch its value from $m-$ to $m+$, which also make the value of A switch from $a-$ (no PSP) to $a+$ (PSP). M is at a higher level than P because M characterizes the whole cell whereas P characterizes functional parts of the cell. Furthermore, M supervenes on P, in the sense that there can be no change in M without a change in P.

[2] Specificity does not require the possibility of complete control. This requirement corresponds to still another notion, which can be analyzed by the requirement that the function from V to L be *surjective*.

The difference between the causal influences of these variables on A can be analyzed in terms of specificity: M is specific for A whereas P is not. M is a two-valued variable that is specific for A according to (S-2). There is a functional association from values of M to values of A. *m*+ is associated with *a*+ and *m*- is associated with *a*-. By contrast, P is a many-valued variable that is *not* specific for A according to (S-P). There is a functional association between values of P and values of A but the function from P to A is not injective. Many different values of P – in fact all those corresponding to the same value of M – are associated with one and the same value of A, either *a*- or *a*+. Say the threshold for triggering a PSP is reached if at least half of the vesicles are merged with the plasma membrane. Then all combinations of values of the *n* components $p_i$ of P, where at least half of the $p_i$ have value $p_i$+ are mapped onto *m*+, and all combinations of values of the $p_1$, where less than half have a value of $p_i$- are mapped onto *m*-.

M is not the only variable representing a property of the cause that causally influences E in the interventionist sense, but it is the most appropriate to mention in an explanation. However, specificity is not relative to any explanatory context. The fact that M is specific for E reflects a fact about the objective dependence of the property represented by E on the property C represented by M, which can be expressed by saying that property C of the cause is causally responsible for property of the effect represented by E.

In an *explanatory* sense, one might say, with Menzies and List, that the choice of M as a cause of E "excludes" the choice of P as a cause of E. However, this way of speaking is misleading because it suggests that the causal influence M→E also *ontologically* excludes the causal influence P→E, which it does not. It is more appropriate to say that both P and M cause E, but that these variables differ as to their specificity with respect to E. This difference explains why it is more appropriate to mention M rather than P in an explanation of E although both are causes of E.

**Reply to objections**

1. The thesis that proportionality (in the sense of a specific degree of determination) can be used to single out higher-level causes has been questioned. Critics (Shapiro and Sober 2012, Franklin-Hall 2014) point out that the "proportionality standard" possesses "no capacity to prefer high-level explanations over low-level ones (or the reverse)" (Franklin-Hall 2014, p. 12) because "there will always be a low-level variable satisfying the letter of the proportionality standard" (Franklin-Hall 2014, p. 13), as it has been expressed by Woodward (2010).

Franklin-Hall argues that there are variables that are *more determinate* than variable $C_x$ (as defined above), which are just as proportional (in Woodward's sense) to the effect E (pecking) than $C_x$. One such variable (the example is not Franklin-Hall's but it is in the same spirit as hers) would be C*, defined as having value *c**+ iff the perceived color is that of monochromatic light with wavelength 550nm and *c**- iff the perceived color is that of light with 300nm. Indeed, C* is a cause of E, and C* also satisfies the proportionality standard as it is spelled out by Woodward. Different values of C* are mapped on different values of E: *c**+ is associated with pecking and *c**- is associated with absence of pecking. However, C* seems intuitively to be too determinate to be the most relevant variable to choose for a causal explanation of E. This refutes Woodward's claim that proportionality correctly characterizes the variable that is the most appropriate for a causal explanation of E.

However, variable C* is no counter-example against the construal of specificity developed above. C* is not specific for E in the sense of our definition (S-2) because C* fails to satisfy clause (2). C* is a variable representing a natural property (the wavelength of light to which a perceiving subject is exposed) that varies within a continuous spectrum. *c**+ corresponds to a part of the spectrum of wavelengths the light can take. However, *the specification of c*- does not fit clause (2)*: *c**- is not defined as corresponding to the complementary part of the spectrum with respect to *c**+.

2. Franklin-Hall examines an analysis she calls "the spirit of proportionality" that results from combining the requirement of proportionality (in Woodward's sense) with a requirement of "exhaustivity" according to which "the cause variable's values collectively exhaust the causal possibility space" (Franklin-Hall, 2014, p. 14). This corresponds to our requirement in clause 2 of (S-2) that the values of the cause variable must together cover all possible values of the causally responsible property. However, she shows that her concept of the spirit of proportionality is not adequate to the purpose of justifying higher-level causation (in the sense of level of determination), because the variable that best fulfills the requirement corresponds to the disjunction of all possible causes of the property represented by the effect variable.

Let us go back once again to Yablo's pigeon. $C_x$ is the variable that is specific for E, representing pecking. However, $C_x$ is not the variable that best fits the requirement of the spirit of proportionality. Franklin-Hall shows that there are variables that fulfill the requirement of exhaustivity better than $C_x$. As an example she offers the variable corresponding to the following disjunction: "the presentation of a red target or provision of food or tickling of the chin or electrical stimulation of the cerebellum (other value: none of the above)" (Franklin-Hall 2014, p. 14). The variable, let us call it $C_{max}$, which takes value $c_{max}$+ in case one of the conditions in this disjunction is fulfilled is indeed just as proportional to E in Woodward's sense as $C_x$, but it fits the standard of the spirit of proportionality better than $C_x$ because the value $c_{max}$+ covers a larger part of the possibility space that includes all possible causes of pecking.

However, $C_{max}$ is not specific for E in the sense of our (S-2) because it does not fit clause (2). Recall that we have raised the question of which variable best represents the property causally responsible for the effect E in the context in which the cause of the event of pecking is already known. Only variables that represent natural properties of that cause event are candidates for being specific for E. We already know that the cause event in terms of the transmission of conserved quantities is the light reaching the pigeon's retina. The interventionist analysis is only used in a second step. Once the cause has been identified as an event, there remains the question as to which of the event's natural properties is causally responsible for the property E of the effect event. The disjunctive predicate constructed by Franklin-Hall does not correspond to any natural property of the light reaching the pigeon's retina.

With the distinction between causation and specific causation, it appears that Menzies and List's claim that there can be "downward exclusion" is after all compatible with the fact that higher-level and downward causal influence are always accompanied by lower-level causal influence.

Downward exclusion in Menzies' and List's sense corresponds to situations in which there is 1) *higher-level influence* $M_1 \rightarrow M_2$ that is *specific* for $M_2$ but there are no lower-level variables $N_1$ and $N_2$ (in the supervenience bases of $M_1$ and $M_2$) for which it would be the case that $N_1 \rightarrow N_2$ is specific, or 2) situations in which there is 1) *downward influence* $M_1 \rightarrow N_2$ that is specific but there is no lower-level variables $N_1$ (in the

supervenience basis of $M_1$) so that $N_1 \rightarrow N_2$ is specific.

## Conclusion

The modified interventionist framework we have sketched makes it possible to justify the claim that statements of higher-level and downward causation are empirically significant. The content of such statements, spelled out in terms of interventionist criteria, differs from the content of the corresponding lower-level causal statements. The empirical truth-conditions of a downward causal statement according to which a higher level variable M influences a lower-level variable E differ from the truth-conditions of the statement that a variable P, which characterizes M's supervenience base, influences E. The absence of higher-level and downward causation can sometimes be justified only inductively but this epistemic problem is no reason to deny that the truth conditions of such statements differ from those of the underlying lower-level statements.

The concept of specificity can be used to distinguish higher-level and downward causal statements in situations where there is both higher-level (or downward) causation and parallel lower-level causation. Specificity makes it possible to explain why it is sometimes more appropriate to mention a higher-level cause in the causal explanation of some variable rather than the underlying lower-level variable. If a higher-level variable M is specific for a given variable E representing a property of the effect event, whereas no lower-level P (in particular physical) variable is specific for E, then it is more appropriate to mention the causal influence of M on E although the underlying P also influences E. The variable that is specific for E represents the property causally responsible for the property of the effect represented by E[3].

## References


- Baumgartner, M. (2009). "Interventionist Causal Exclusion and Non-Reductive Physicalism". *International Studies in the Philosophy of Science* 23 (2):161-178.
- Baumgartner, M. (2010). "Interventionism and Epiphenomenalism"**.** *Canadian Journal of Philosophy* 40: 359-383
- Baumgartner, M. (2013), Rendering Interventionism and Non-Reductive Physicalism Compatible, *Dialectica* 67: 1-27.
- Baumgartner, M. and A. Gebharter (2015), Constitutive relevance, mutual manipulability, and fat-handedess, *British Journal for the Philosophy of Science*, online 20 February, 2015 : 1-26 ; doi: 10.1093/bjps/axv003.
- Craver, C. F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford, Oxford University Press.
- Cuijpers P., van Straten A., van Oppen P., Andersson G. (2008), Are psychological and pharmacologic interventions equally effective in the treatment of adult depressive disorders? A meta-analysis of comparative studies. *J Clin Psychiatry*. 2008 Nov; 69(11), p. 1675-85.
- Franklin-Hall, L.R. (2014), High-Level Explanation and the Interventionist's 'Variables Problem', *British Journal for the Philosophy of Science*, p. 1-25.



[3] Versions of this paper have been presented at conferences in Montreal, Macerata and Glasgow. I thank my auditors for their critical remarks.

- Kandel E. and S. Siegelbaum (2000), Signaling at the Nerve-Muscle Synapse: Directly Gated Transmission, chap. 11 in Kandel, Schwartz, Jessell, *Principles of Neural Science*.
- Kalia, M. (2005), Neurobiological basis of depression: an update. *Metabolism* 54(5 Suppl 1), p. 24-7.
- Kim, J. (1998), *Mind in a Physical World*, MIT Press.
- Kistler, M. (2006), *Causation and Laws of Nature*. London, Routledge.
- Kistler, M. (2013), The Interventionist Account of Causation and Non-causal Association Laws, *Erkenntnis*. 78: 65–84.
- Kistler, M. (2014), Analysing Causation in Light of Intuitions, Causal Statements, and Science, in B. Copley, F. Martin (eds.), *Causation in Grammatical Structures*, Oxford University Press, 2014, p 76-99.
- List C. and P. Menzies (2009), Non-Reductive Physicalism and the Limits of the Exclusion Principle, *Journal of Philosophy* 106, p. 475-502.
- Marcellesi, A. (2010), L'interventionnisme permet-il la causalité "descendante" ?, *Igitur* 2, p. 1-11.
- Menzies P. and C. List (2010), The Causal Autonomy of the Special Sciences, in C. Macdonald and G. Macdonald, *Emergence in Mind,* OUP, p. 108-128.
- Raatikainen, P. (2010). "Causation, Exclusion, and the Special Sciences". *Erkenntnis* 73, p. 349-363.
- Shapiro, L. and Sober E. (2007), Epiphenomenalism. The Dos and Don'ts. In G. Wolter and P. Machamer (eds.), *Thinking about Causes: From Greek Philosophy to Modern Physics*, Pittsburgh, Pittsburgh University Press, p. 235-264.
- Shapiro, L. and Sober, E. (2012), Against Proportionality, *Analysis* 72, p. 89-93.
- Woodward, J. (2003), *Making Things Happen*, Oxford University Press.
- Woodward, J. (2008), Mental Causation and Neural Mechanisms, in J. Hohwy and J. Kallestrup, *Being Reduced*, Oxford, Oxford University Press, p. 218-262.
- Woodward, J. (2008, b), Cause and Explanation in Psychiatry, in K. Kendler and J. Parnas (eds.), *Philosophical Issues in Psychiatry: Explanation, Phenomenology and Nosology*. Johns Hopkins University Press, p. 132-183.
- Woodward (2010), Causation in biology : stability, specificity, and the choice of levels of explanation, *Biology and Philosophy* 25, p. 287-318.
- Woodward, J. (2014), Interventionism and Causal Exclusion, *Philosophy and Phenomenological Research*, online 7 april 2014; http://philsci-archive.pitt.edu/8651.
- Woodward, J. (forthcoming), Intervening in the Exclusion Argument, in H. Beebee, C. Hitchcock, H. Price (eds.), *Making a Difference*.
- Yablo, Stephen (1992), Mental Causation, *Philosophical Review* 101, p. 245-280.
- Yablo, Stephen (1997), Wide Causation, *Philosophical Perspectives* 11: Mind, Cause and World, p. 251-281.